SORT 28 (2) July-December 2004, 201-214

Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes

Célestin C. Kokonendji, Simplice Dossou-Gbété*, Clarice G.B. Demétrio**

* *University of Pau-LMA, Pau, France,* ** *University of São Paulo*

Abstract

In this paper we investigate two classes of exponential dispersion models (EDMs) for overdispersed count data with respect to the Poisson distribution. The first is a class of Poisson mixture with positive Tweedie mixing distributions. As an approximation (in terms of unit variance function) of the first, the second is a new class of EDMs characterized by their unit variance functions of the form $\mu + \mu^p$, where p is a real index related to a precise model. These two classes provide some alternatives to the negative binomial distribution ($p = 2$) which is classically used in the framework of regression models for count data when overdispersion results in a lack of fit of the Poisson regression model. Some properties are then studied and the practical usefulness is also discussed.

MSC: 62E15, 62E17, 60E07

Keywords: Negative binomial distribution, overdispersion, Poisson mixture, Tweedie family, unit variance function

1 Introduction

The Poisson distribution is well-known to be the classical distribution for count data, but it has only one parameter and its variance is equal to the mean. Since the index of dispersion (i.e. the variance divided by the mean) of Poisson is one, this makes it inadequate for fitting overdispersed count data (e.g. Castillo and Pérez-Casany, 2004), and raises the question of whether an appropriate two-parameter distribution such as the negative binomial should be used routinely for analysing overdispersed count data. The same problem occurs in the framework of regression models for count data

* *Address for correspondence:* C.C. Kokonendji. Université de Pau et des Pays de l'Adour. Laboratoire de Mathématiques Appliquées. CNRS FRE 2570-Département STID. Avenue de l'Université. 64000 Pau, France. Phone: +33 559 407 145. Fax: +33 559 407 140.

** University of São Paulo - ESALQ, Piracicaba - SP, Brazil.

Received: January 2004
Accepted: September 2004

(McCullagh and Nelder, 1989), when the Poisson distribution does not fit well, and the observed dispersion is greater than that predicted by the standard distribution.

It is well known that negative binomial can be understood as a Poisson mixture with gamma mixing distribution, taking into account the heterogeneity in the population. Hougaard *et al* (1997) have considered a large family of mixture distributions, including the Poisson-inverse Gaussian distribution, to improve significantly the fitness to certain data. We will call the *Poisson-Tweedie* class a completed set of these distributions that we must point out the exact form of its associated “unit variance function” (a term to be made precise). Otherwise, Hinde and Demétrio (1998, page 14) propose for overdispersed count data the use of the unit variance function

$$V_p^{HD}(\mu) = \mu + \mu^p, \quad \mu \in M_p^{HD} \subseteq \mathbb{R}, \tag{1}$$

where $p \in \mathbb{R}$ fixed, which is also an alternative to negative binomial unit variance function obtained with $p = 2$ and includes the strict arcsine distribution with $p = 3$ (Kokonendji and Khoudar, 2004). We here call the *Hinde-Demétrio* class the set of all distributions associated to (1). The aim of this work is to provide a complete identification of both the Poisson-Tweedie and the Hinde-Demétrio classes from their unit variance functions. These classes are sets of two-parameter distributions with an additional index parameter p allowing to identify an appropriate family of these distributions.

In Section 2, we review some basic properties of the general models, called “exponential dispersion models” and, in particular, we present the *Tweedie* class with unit variance function μ^p. In Section 3, we describe the possible Poisson mixture distributions with a Tweedie for obtaining the Poisson-Tweedie class: unit variance function and probabilities are given. In Section 4, we first classify the Hinde-Demétrio class (1) and we then compare it to the Poisson-Tweedie class. We stress that there is no intersection between the Hinde-Demétrio class ($\mu + \mu^p$) and the Tweedie class (μ^p), except for $p = 2$. Section 5 is devoted to concluding remarks and the problem of statistical inference for p to select the adequate model in these classes.

2 Exponential dispersion models

Exponential dispersion models (Jørgensen, 1997) are important in statistical modelling. They have a number of important mathematical properties, which are relevant in practice. They include several well-known families of distributions as special cases, giving a convenient general framework. Generalized linear models (McCullagh and Nelder, 1989) are based on these families of distributions.

Let ν be a σ-finite positive measure on the real line $\mathbb{R}$ (not necessarily a probability) and define the cumulant function K by

$$K(\theta) = \ln \int_{\mathbb{R}} \exp(\theta x)\nu(dx)$$

on its (canonical parameter) domain $\Theta = \{\theta \in \mathbb{R} : K(\theta) < \infty\}$. Assume that both ν and Θ are not degenerate (i.e., ν is not concentrated at one point and the interior of Θ is not empty), then the set of the probability measures $P(\theta;\nu)(dx) = \exp\{\theta x - K(\theta)\}\nu(dx)$ defined for all θ in $\Theta = \Theta(\nu)$ represents a *natural exponential family* (NEF) generated by ν and denoted $F = F(\nu) = \{P(\theta;\nu); \theta \in \Theta(\nu)\}$; see Chapter 54 of Kotz *et al* (2000). Given a NEF, we define the set Λ of reals $\lambda > 0$ such that $\lambda K(\theta)$ is also the cumulant function for some measure ν_λ. For fixed $\lambda \in \Lambda$, the NEF $F_\lambda = F(\nu_\lambda)$ generated by ν_λ is then $\exp\{\theta x - \lambda K(\theta)\}\nu_\lambda(dx)$, for $\theta \in \Theta$. This family of distributions, denoted $\mathcal{ED}(\theta,\lambda)$ for $(\theta,\lambda) \in \Theta \times \Lambda$, is called the *exponential dispersion model* (EDM) generated by ν (or ν_λ for improper notation); and λ can be called the dispersion parameter. Its density or mass function with respect to some measure η can be written as

$$C(x;\lambda)\exp\{\theta x - \lambda K(\theta)\}, \quad x \in S \subseteq \mathbb{R}, \tag{2}$$

where $\nu_\lambda(dx) = C(x;\lambda)\eta(dx)$. Note here that $\mathcal{ED}(\theta,\lambda)$ defined by (2) is the additive version of EDM. The reproductive version of $X \sim \mathcal{ED}(\theta,\lambda)$ is given by $Z = X/\lambda$. However, additive EDMs turn out to be important for discrete data because many usefull families of discrete distributions have this form. Any EDM satisfies $\mathcal{ED}(\theta,\lambda_1) * \mathcal{ED}(\theta,\lambda_2) = \mathcal{ED}(\theta,\lambda_1+\lambda_2)$, so the family is closed under convolution and $\{1,2,\cdots\} \subseteq \Lambda$. Also the model is infinitely divisible if and only if $\Lambda = (0,\infty)$.

In the interior of Θ, denoted intΘ, the cumulant function $\theta \mapsto K(\theta)$ is strictly convex. Then the expectation and variance of $X \sim \mathcal{ED}(\theta,\lambda)$ are

$$\mathbb{E}(X) = \lambda K'(\theta) \quad \text{and} \quad \mathrm{Var}(X) = \lambda K''(\theta), \tag{3}$$

where $K'(\theta)$ and $K''(\theta)$ are, respectively, the first and second derivatives of K at the point θ. From (3) with $\lambda = 1$, the characterizing function V defined on the domain $M = K'(\mathrm{int}\Theta)$ such that

$$K''(\theta) = V\{K'(\theta)\}$$

is called *unit variance function*. We also have $V(\mu) = 1/\psi'(\mu)$, for $\mu \in M$, where $\psi = (K')^{-1}$ is the inverse function of K'. Note that M depends only on the family $F = \{\mathcal{ED}(\theta,1) : \theta \in \Theta\}$, and not on the choice of the generating measure ν of F. If $M = \Omega$, where Ω denotes the interior of the convex hull of the support S of F, the family F is said to be *steep*. From here to the end, an EDM is always assumed to be steep. The role of the unit variance function in data fitting should be to identify an appropriate EDM of distributions, if any. The reparametrization by unit mean $\mu = K'(\theta) = \mu(\theta)$ allows us to write the EDM as follows: $\{\mathcal{ED}(\mu(\theta),\lambda); \mu(\theta) \in M, \lambda \in \Lambda\} \equiv EDM(\mu,\lambda)$. It is sometimes considered the reparametrization of the EDM by the mean $m = \mathbb{E}(X) = \lambda K'(\theta) = m(\lambda,\theta)$ instead of the unit mean $\mu = \mu(\theta)$. From (3), the unit variance function V leads to the variance $V_\lambda = \mathrm{Var}(X)$ of $X \sim \mathcal{ED}(\theta,\lambda)$ in terms of m, called *variance function* and expressed as follows: $V_\lambda(m) = \lambda V(m/\lambda)$, for all $m \in \lambda M$. For discrete overdispersed EDM compared to the Poisson distribution we have

$$V(\mu) > \mu, \quad \mu > 0, \tag{4}$$

where $V(\mu) = \mu$ is the unit variance function of the Poisson model (e.g. Jourdan and Kokonendji, 2002).

A complete description of the EDMs with power unit variance functions

$$V_p^T(\mu) = \mu^p, \quad p \in (-\infty, 0] \cup [1, \infty), \tag{5}$$

is given by Jørgensen (1997) where, for $p \to \infty$ the corresponding unit variance function takes the exponential form $V_\infty^T(\mu) = \exp(\beta\mu)$, $\beta \neq 0$. This class, called the *Tweedie class*, was introduced by Tweedie (1984). It is also convenient to introduce the index parameter α of stable distribution, connected to p by the following relation:

$$(p-1)(1-\alpha) = 1. \tag{6}$$

According to the above notations, we can denote by $\mathcal{T}_p(\theta, \lambda)$ any distribution of this class where $\lambda \in (0, \infty) = \Lambda$ for all p of (5), $\mu \in M_p = K'(\mathrm{int}\Theta_p)$ and $\theta \in \Theta_p$ with

$$\Theta_p = \begin{cases} \mathbb{R} & \text{for} \quad p = 0, 1 \\ [0, \infty) & \text{for} \quad p < 0 \ \text{ or } 0 < p < 1 \\ (-\infty, 0) & \text{for} \quad 1 < p \leq 2 \ \text{ or } p \to \infty \\ (-\infty, 0] & \text{for} \quad 2 < p < \infty. \end{cases} \tag{7}$$

Thus, for $s \in \Theta_p - \theta$ and $X \sim \mathcal{T}_p(\theta, \lambda)$, the Laplace transform $\mathbb{E}(e^{sX})$ is

$$G_p(s; \theta, \lambda) = \begin{cases} \exp\left\{ \dfrac{\lambda[(1-p)\theta]^\alpha}{(2-p)} [(1 + s/\theta)^\alpha - 1] \right\} & \text{for} \quad p \neq 1, 2 \\ (1 + s/\theta)^{-\lambda} & \text{for} \quad p = 2 \\ \exp\{\lambda e^\theta (e^s - 1)\} & \text{for} \quad p = 1. \end{cases} \tag{8}$$

As shown in Table 1, the Tweedie class $T = \{TM_p(\mu, \lambda); p \in \mathbb{R}\}$ includes several well-known families of distributions amongst which one may be the inverse-Gaussian model $TM_3(\mu, \lambda)$ and the noncentral gamma model $TM_{3/2}(\mu, \lambda)$ of zero shape (respectively, a special case of positive stable and compound Poisson families). The compound Poisson $(1 < p < 2)$ is also called Poisson-gamma; it can be represented as the Poisson random sum of independent gamma random variables (and it has a mass at zero but otherwise has a continuous positive distribution). Observe, however, that the extreme stable distributions $(p < 0)$ are not steep and only the Poisson distribution $(p = 1)$ is discrete.

Table 1: *Summary of Tweedie EDMs (Jørgensen, 1997).*

Distribution	p	α	M	S
Extreme stable	$p < 0$	$1 < \alpha < 2$	$(0, \infty)$	$\mathbb{R}$
Normal	$p = 0$	$\alpha = 2$	$\mathbb{R}$	$\mathbb{R}$
[Do not exist]	$0 < p < 1$	$2 < \alpha < \infty$		
Poisson	$p = 1$	$\alpha \to -\infty$	$(0, \infty)$	$\mathbb{N}$
Compound Poisson	$1 < p < 2$	$\alpha < 0$	$(0, \infty)$	$(0, \infty)$
Gamma	$p = 2$	$\alpha = 0$	$(0, \infty)$	$(0, \infty)$
Positive stable	$p > 2$	$0 < \alpha < 1$	$(0, \infty)$	$[0, \infty)$
Extreme stable	$p \to \infty$	$\alpha = 1$	$\mathbb{R}$	$\mathbb{R}$

3 Poisson-Tweedie EDMs

Let X be a non-negative random variable following $\mathcal{T}_p(\theta, \lambda)$. If a discrete random variable Y is such that the conditional distribution of Y given X is Poisson with mean X, then the EDM generated by the distribution of Y is of the Poisson-Tweedie class. We can also use the following notations $\mathcal{PT}_p(\theta, \lambda)$ to denote the distribution of Y and $PTM_p(\mu, \lambda)$ for the corresponding EDM. Hence for $p \geq 1$, the individual probabilities of $Y \sim \mathcal{PT}_p(\theta, \lambda)$ are

$$\Pr(Y = y) = \int_0^\infty \frac{e^{-x} x^y}{y!} \mathcal{T}_p(\theta, \lambda)(dx), \quad y = 0, 1, \cdots. \tag{9}$$

Proposition 1 *(Hougaard et al., 1997) Let $Y \sim \mathcal{PT}_p(\theta, \lambda)$ defined by (9), where $\theta \in \Theta_p$ given by (7) and $\lambda > 0$ for fixed $p \geq 1$ or $\alpha \in [-\infty, 1)$ from (6). We have the following properties: (i) If $Y_1, \cdots, Y_n$ are independent, with $Y_i \sim \mathcal{PT}_p(\theta, \lambda_i)$, then $Y_1 + \cdots + Y_n$ follows $\mathcal{PT}_p(\theta, \lambda_1 + \cdots + \lambda_n)$. The distribution $\mathcal{PT}_p(\theta, \lambda)$ is infinitely divisible. (ii) The distribution $\mathcal{PT}_p(\theta, \lambda)$ is unimodal for $p \geq 2$. (iii) The Laplace transform of $Y \sim \mathcal{PT}_p(\theta, \lambda)$ is*

$$\mathbb{E}(e^{\omega Y}) = \begin{cases} \exp\left\{\dfrac{\lambda}{2-p}[\{(1-p)(e^\omega - 1 + \theta)\}^\alpha - \{(1-p)\theta\}^\alpha]\right\} & \text{for} \quad p \neq 1, 2 \\ [(e^\omega - 1 + \theta)/\theta]^{-\lambda} & \text{for} \quad p = 2 \\ \exp\left\{\lambda[\exp(e^\omega - 1 + \theta) - e^\theta]\right\} & \text{for} \quad p = 1, \end{cases} \tag{10}$$

for $\omega \in \Theta_p - \theta$. For $p = 1$, it is a Neyman type A distribution; for $p = 2$, it is a negative binomial distribution; and, for $p = 3$, it is the Sichel or Poisson-inverse Gaussian distribution (e.g. Willmot, 1987).

Proposition 2 *With the asumptions of Proposition 1, the unit variance function of the model $PTM_p(\mu,\lambda)$ generated by $Y \sim \mathcal{PT}_p(\theta,1)$ is exactly*

$$V_p^{PT}(\mu) = \mu + \mu^p \exp\{(2-p)\Phi_p(\mu)\}, \quad \mu > 0, \tag{11}$$

where $\Phi_p(\mu)$, generally implicit, denotes the inverse of the increasing function $\omega \mapsto d\{\ln \mathbb{E}(e^{\omega Y})\}/d\omega$.

Proof. Let $K(\omega) = \ln \mathbb{E}(e^{\omega Y})$ for $Y \sim \mathcal{PT}_p(\theta,1)$. From Proposition 1 (*iii*) with $\lambda = 1$ and using (6) to simplify, the first derivative of $K(\omega)$ is

$$\mu = K'(\omega) = \begin{cases} e^{\omega}[(1-p)(e^{\omega}-1+\theta)]^{\alpha-1} & \text{for} \quad p \neq 1,2 \\ -e^{\omega}(e^{\omega}-1+\theta)^{-1} & \text{for} \quad p = 2 \\ e^{\omega}\exp\{e^{\omega}-1+\theta\} & \text{for} \quad p = 1, \end{cases}$$

and the second derivative of $K(\omega)$ may be expressed as follows:

$$V_p^{PT}(\mu) = K''(\omega) = \begin{cases} K'(\omega) + e^{2\omega}[(1-p)(e^{\omega}-1+\theta)]^{\alpha-2} & \text{for} \quad p \neq 1,2 \\ K'(\omega) + [K'(\omega)]^2 & \text{for} \quad p = 2 \\ K'(\omega) + e^{\omega}K'(\omega) & \text{for} \quad p = 1. \end{cases}$$

For $p \neq 1,2$, we can also write $K''(\omega) = K'(\omega) + e^{2\omega}[K'(\omega)/e^{\omega}]^p$ and the expression given in (11) is easily obtained. □

The Poisson-Tweedie EDMs are summarized in Table 2, that we can divide in three parts with respect to p: $1 < p < 2$, $2 < p < \infty$ and $p \in \{1,2,\infty\}$.

Table 2: *Summary of Poisson-Tweedie EDMs.*

Distribution	p	α	M	S
[Do not define]	$p < 1$	$1 < \alpha < \infty$		
Neyman type A	$p = 1$	$\alpha \to -\infty$	$(0,\infty)$	$\mathbb{N}$
Poisson-compound Poisson	$1 < p < 2$	$\alpha < 0$	$(0,\infty)$	$\mathbb{N}$
Negative binomial	$p = 2$	$\alpha = 0$	$(0,\infty)$	$\mathbb{N}$
Poisson-positive stable	$p > 2$	$0 < \alpha < 1$	$(0,\infty)$	$\mathbb{N}$
Poisson	$p \to \infty$	$\alpha = 1$	$(0,\infty)$	$\mathbb{N}$

For $p = 1$ or $\alpha \to -\infty$ which is not studied by Hougaard *et al*. (1997), we can refer to Johnson *et al.* (1992; pages 368-) to obtain some properties on the Neyman type A distribution, which is therefore both a Poisson mixture of Poisson distributions, and also a Poisson-stopped sum of Poisson distributions.

Note that we consider in this paper only the strictly mixed Poisson distributions ($p \geq 1$). In fact, it is not possible to mix a Poisson with a Tweedie distribution $\mathcal{T}_p(\theta,\lambda)$ for $p \leq 0$ because it can be negative but (9) can be seen as a purely formal operation. For $p = 0$ we refer to Kemp and Kemp (1966) who show that, if X follows a normal distribution $\mathcal{T}_0(\theta,\lambda)$ with mean $\mu = \mu(\theta)$ and standard deviation $\sigma = \sigma(\lambda)$ such that $\mu \geq \sigma^2$, the corresponding mixed Poisson distribution is the Hermite distribution (Johnson *et al.*, 1992; pages 357-364).

To conclude this section, we explicit the probability mass functions (2) of all the Poisson-Tweedie EDMs generated by any distribution of $Y \sim \mathcal{PT}_p(\theta_0, \lambda)$ for fixed $p \in [1, \infty)$ and $\theta_0 \in \overline{\Theta}_p$ the closure of Θ_p given in (7). Indeed, one has

$$C_{p,\theta_0}(y;\lambda)\exp\{\omega y - \lambda K_{p,\theta_0}(\omega)\}, \quad y = 0, 1, 2, \cdots, \tag{12}$$

where $\omega \in \Theta_p - \theta_0$ is the canonical parameter and $\lambda > 0$ is the dispersion parameter such that, respectively by (9) and (10),

$$C_{p,\theta_0}(y;\lambda) = \frac{1}{y!}\mathbb{E}(e^{-X}X^y) = \frac{1}{y!}\frac{\partial^y G_p(s;\theta_0,\lambda)}{\partial s^y}|_{s=-1} \text{ for } X \sim \mathcal{T}_p(\theta_0,\lambda) \tag{13}$$

and $K_{p,\theta_0}(\omega) = \lambda^{-1}\ln\mathbb{E}(e^{\omega Y})$ for $Y \sim \mathcal{PT}_p(\theta_0, \lambda)$. Note that, in practice, we can use $\theta_0 = 0$ which is here defined for all $p \in [1, \infty)$ with the convenience: $0^\lambda = 1$, for any $\lambda > 0$ (only for $p = 2$). To clarify completely (12), we must point out $C_{p,\theta_0}(y;\lambda)$ by the following proposition.

Proposition 3 *Let $p \in [1, \infty)$ be fixed and let $\theta_0 \in \overline{\Theta}_p$ given by (7). Then, for all $\lambda > 0$ and $y \in \mathbb{N}$, the expression of $C_{p,\theta_0}(y;\lambda)$ in (12) is,*

- for $p = 1$ or $\alpha \to -\infty$,

$$C_{1,\theta_0}(y;\lambda) = \begin{cases} \exp\{\lambda e^{\theta_0-1}(1-e)\} & \text{for} \quad y = 0 \\ \dfrac{1}{y!}C_{1,\theta_0}(0;\lambda)\sum\limits_{k=1}^{y} a_{y,k}(\lambda e^{\theta_0-1})^k & \text{for} \quad y = 1, 2, \cdots, \end{cases}$$

with $a_{y,y} = a_{y,1} = 1$ and $a_{y,k} = a_{y-1,k-1} + ka_{y-1,k}$;

– for $p = 2$ or $\alpha = 0$, $C_{2,\theta_0}(y;\lambda) = \left(\dfrac{\theta_0}{\theta_0 - 1}\right)^\lambda \dfrac{\Gamma(y+\lambda)}{y!\Gamma(\lambda)}\left(\dfrac{1}{1-\theta_0}\right)^y$*;*

– for $p \in (1, 2) \cup (2, \infty)$ or $\alpha \in (-\infty, 0) \cup (0, 1)$,

$$C_{p,\theta_0}(y;\lambda) = \begin{cases} \exp\left\{\dfrac{\lambda(\alpha-1)}{\alpha}[\{(1-\theta_0)/(1-\alpha)\}^\alpha - \{(-\theta_0)/(1-\alpha)\}^\alpha]\right\} & \text{for} \quad y = 0 \\ \dfrac{1}{y!}C_{p,\theta_0}(0;\lambda)\sum\limits_{k=1}^{y} a_{y,k}(\alpha)\lambda^k\{(1-\theta_0)/(1-\alpha)\}^{k\alpha-y} & \text{for}\, y = 1, 2, \cdots, \end{cases}$$

with $a_{y,y}(\alpha) = 1$, $a_{y,1}(\alpha) = \Gamma(y-\alpha)/[(1-\alpha)^{y-2}\Gamma(2-\alpha)]$ and $a_{y,k}(\alpha) = a_{y-1,k-1}(\alpha) + [(y-1-k\alpha)/(1-\alpha)]a_{y-1,k}(\alpha)$.

Proof. Following Hougaard *et al.* (1997), we show only for $p = 1$ or $\alpha \to -\infty$. From (13), it suffices to check the recursive formula of derivatives of Laplace transform (8), which is $\partial^y G_1(s;\theta_0,\lambda)/\partial s^y = G_1(s;\theta_0,\lambda)\sum_{k=1}^{y} a_{y,k}(\lambda e^{\theta_0+s})^k$. □

4 Hinde-Demétrio EDMs

We now characterize the Hinde-Demétrio class which is the set of EDMs with unit variance function of the "simple" form (1) and, then, we compare it to the Poisson-Tweedie class (11).

Theorem 4 *Let $p > 1$. Then there exists a NEF $F_{p,1}$ such that $M_{F_{p,1}} = (0, \infty)$ and $V_{F_{p,1}}(\mu) = \mu + \mu^p$. Furthermore, the NEF $F_{p,1}$ is infinitely divisible (with bounded Lévy measure). More precisely, denote $a = 1/(p-1)$ similarly to the interchangeable relation (6) and consider the positive measure*

$$\nu_p = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)}{k!} \frac{1}{1+k(p-1)} \delta_{1+k(p-1)}.$$

Then, for all $\lambda > 0$, $\nu_{p,\lambda} = \exp\{\lambda \nu_p\} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \nu_p^{*n}$ generates a NEF $F_{p,\lambda}$ with variance function

$$V_{F_{p,\lambda}}(m) = m + \lambda^{1-p} m^p \tag{14}$$

and $F_{p,\lambda}$ is concentrated on the additive semigroup $\mathbb{N} + p\mathbb{N}$.

Proof. It is standard by checking that we have exactly

$$K'_{\nu_{p,1}}(\theta) = \frac{e^{\theta}}{[1 - e^{\theta(p-1)}]^a} = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)}{k!} e^{\theta[1+k(p-1)]}.$$

□

For $p > 1$, we also denote by $\mathcal{HD}_p(\theta, \lambda)$ any distribution of the EDM $HDM_p(\mu, \lambda)$ corresponding to (1) or (14) with $\lambda > 0$ and $\theta \in \Theta_p^* \subseteq \mathbb{R}$. As a probabilistic interpretation of $Y \sim \mathcal{HD}_p(\theta, \lambda)$, let X be a random variable associated to $\nu_p \equiv \nu_{p,\theta}$ of Theorem 4 (up to normalizing constant) and, let N_t be a standard Poisson process on the interval $(0, t]$ ($N_0 = 0$) with rate λ (i.e. $N_t \sim \mathcal{P}(\lambda t)$) and supposed to be independent of X. From the Laplace transform of

$$Y_t = \sum_{i=1}^{N_t} X_i = X_1 + \cdots + X_{N_t},$$

where the X_i are independent and identically distributed as X, it is easy to see that $Y = Y_1$ by fixing the time to $t = 1$.

Theorem 5 *Let $p \in \mathbb{R}$. The function (1): $\mu \mapsto V_p^{HD}(\mu) = \mu + \mu^p$, defined on a suitable domain M_p^{HD} corresponds to a unit variance function of discrete (steep) EDM when*

$$p \in \{0\} \cup [1, \infty),$$

with $M_0^{HD} = (-1, \infty)$ and $M_p^{HD} = (0, \infty)$ for $p \geq 1$; and the domain Θ_p of the canonical parameter is given by (7). Moreover, if $p = 0$ the model $HDM_0(\mu, \lambda)$ is a positive-translated Poisson; if $p = 1$ the model $HDM_1(\mu, \lambda)$ is a scaled Poisson; if $p = 2$

the model $HDM_2(\mu,\lambda)$ *is negative binomial; if* $p = 3$ *the model* $HDM_3(\mu,\lambda)$ *is strict arcsine (Kokonendji and Khoudar, 2004); and, if* $p \neq 0,1,2,3$ *the model* $HDM_p(\mu,\lambda)$ *is deduced from Theorem 4.*

Before giving the proof, let us observe that, from (4), the Hinde-Demétrio class (1) is the set of overdispersed EDMs compared with the Poisson distribution, as well as the Poisson-Tweedie class (11) (see also Feller, 1943). The next proposition provides comparable behaviours of these two classes. Indeed, only negative binomial model $HDM_2(\mu,\lambda)$ is interpreted as $PTM_2(\mu,\lambda)$; and, for fixed $p \geq 1$ and $\lambda > 0$, each $HDM_p(\mu,\lambda)$ can be "approximated" by $PTM_p(\mu,\lambda)$ for large μ. This approximation must be understand in terms of their unit variance functions.

Proposition 6 *Let* $HD = \{HDM_p(\mu,\lambda); p \in \mathbb{R}\}$ *be the Hinde-Demétrio class (1) and* $PT = \{PTM_p(\mu,\lambda); p \in \mathbb{R}\}$ *the Poisson-Tweedie class (11). Then: (i)* $HD \cap PT = \{HDM_2(\mu,\lambda) = PTM_2(\mu,\lambda)\}$*; (ii) For fixed* $p \geq 1$, $V_p^{PT}(\mu) \sim V_p^{HD}(\mu)$ *as* $\mu \to \infty$.

For the proof of Theorem 5 we need the two following lemmas. The first is an "impossibility criterion" to exclude case $0 < p < 1$, and the second is related to the steepness.

Lemma 7 *There are no EDM with* $M = (0,\infty)$ *and unit variance function* $V(\mu) \sim \mu^\gamma$ *as* $\mu \to 0$ *for* $\gamma \in (0,1)$.

Proof. If $V(\mu) \sim \mu^\gamma$ as $\mu \to 0$, then $\theta = \psi(\mu) = \theta_0 + \int_0^\mu (V(t))^{-1}dt$ is left-bounded. Now, $V(\mu) \to 0$ as $\mu \to 0$ implies that the generator ν is concentrated on $[0,\infty)$ (e.g. Letac and Mora, 1990). Hence, the domain $\Theta(\nu)$ is not left-bounded, which yields a contradiction. □

Lemma 8 *(Jørgensen, 1997; page 58) Let* $F = \{\mathcal{ED}(\theta,1)\,;\theta \in \Theta\}$ *be a NEF with variance function* V *on* M *and support* S*. If* $\inf S = 0$*, then: (i)* $\inf M = 0$*; (ii)* $\lim_{\mu\to 0} V(\mu) = 0$*; (iii)* $\lim_{\mu\to 0} V(\mu)/\mu = c$*, where* $c = \inf\{S \setminus \{0\}\}$.

Note that $c = 0$ for continuous distributions and $c = 1$ for discrete integer-valued distributions.

Proof of Theorem 5: Since V_p^{HD} must be an analytic positive function on the domain $M_p^{HD} = (a,\infty)$, we have that V_p^{HD} has no zeroes in (a,∞) and $V_p^{HD}(a) = 0$ [this is a consequence of Theorem 3.1 of Letac and Mora, 1990]. Thus, we have

$$M_p^{HD} = \begin{cases} (0,\infty) & \text{for} \quad p \neq 0 \\ (-1,\infty) & \text{for} \quad p = 0. \end{cases} \tag{15}$$

Solving $\psi'(\mu) = 1/V_p^{HD}(\mu) = 1/(\mu + \mu^p)$ and ignoring the arbitrary constants in the solutions,

$$\psi(\mu) = \begin{cases} \ln(\mu) - (p-1)^{-1}\ln(1+\mu^{p-1}) & \text{for} \quad p \neq 0,1 \\ \ln\sqrt{\mu} & \text{for} \quad p = 1 \\ \ln(1+\mu) & \text{for} \quad p = 0. \end{cases} \tag{16}$$

We now examine the different situations of $p \in \mathbb{R}$ in (1) from (15).

– Consider case $p \in \{0\} \cup [1,\infty)$. Let $\theta = \psi(\mu)$ be the canonical link function given by (16), then we find first $\mu = K'(\theta)$ and, hence,

$$K(\theta) = \begin{cases} e^{\theta} - \theta & \text{for} \quad p = 0 \\ e^{2\theta}/2 & \text{for} \quad p = 1 \\ -\ln(1-e^{\theta}) & \text{for} \quad p = 2 \\ \arcsin e^{\theta} & \text{for} \quad p = 3 \\ \sum_{k=0}^{\infty} \dfrac{\Gamma[k+1/(p-1)]}{k!\Gamma[1/(p-1)]} \dfrac{\exp\{\theta[1+k(p-1)]\}}{1+k(p-1)} & \text{for} \quad p \neq 0,1,2,3, \end{cases}$$

for $\theta \in \Theta$, where the interior of Θ is obtained by using (15) and (16):

$$\mathrm{int}\Theta = \begin{cases} \mathbb{R} & \text{for} \quad p = 0,1 \\ (0,\infty) & \text{for} \quad p < 0 \text{ or } 0 < p < 1 \\ (-\infty,0) & \text{for} \quad 1 < p < \infty. \end{cases} \tag{17}$$

Since $K(\theta)$ is analytic, the domain Θ defined from its interior (17) coincides to Θ_p given by (7). Thus, for each $p \in \{0\} \cup [1,\infty)$, we define a discrete generator of the corresponding (steep) EDM with unit variance function (1).

– Case $0 < p < 1$ is excluded by Lemma 7.

– Finally, let us exclude case $p < 0$ by Lemma 8. Indeed, it suffices to observe that $M_p^{HD} = (0,\infty)$ from (15) and $\lim_{\mu\to 0} V_p^{HD}(\mu)/\mu = \lim_{\mu\to 0}(1+\mu^{p-1}) = \infty$. The proof of Theorem 5 is now complete. □

5 Concluding remarks

Here we have two classes of two parameter distributions which could be used as models for analysing overdispersed count data. It was showed that both are EDMs with general unit variance functions indexed by a third parameter p. A common member of both families is the negative binomial family when $p = 2$. However, the probability mass functions (2) of $HDM_p(\mu,\lambda)$ are generally not easy to calculate (except for $p \in \{0,1,2,3\}$) whereas for $PTM_p(\mu,\lambda)$ are given explicitly by (12). When μ is large, Proposition 6 (*ii*) allows the use of the Poisson-Tweedie model as well as the Hinde-Demétrio model since the variance functions are equivalent.

For models with covariates, let Y be a count response variable and let $\mathbf{x}$ be an associated $d \times 1$ vector of covariates with a vector β of unknown regression coefficients, the relation between the mean $m = \mathbb{E}(Y) = m(\mathbf{x};\beta)$ of the distribution and the linear part

$\mathbf{x}^T\boldsymbol{\beta}$ being made through a link function (McCullagh and Nelder, 1989). For both EDMs the cumulant functions are given in explicit form. This allows to compute deviances and then to use maximum likelihood method for the estimation of the parameters.

When the Poisson-Tweedie models $PTM_p(\mu, \lambda)$ or the Hinde-Demétrio models $HDM_p(\mu, \lambda)$ are used, one of the problems for statistical inference is the index parameter p of the adequate distribution. A profile estimate of p is recommended in the general situation; see Hougaard *et al.* (1997) for Poisson-Tweedie mixture. In Hinde-Demétrio models, we can start by the moment estimate of p. Indeed, if $\underline{y} = (y_1, \cdots, y_n)$ is an n-independent identically distributed observation from $\mathcal{HD}_p(\mu, \lambda)$ such that the overdispersion condition $s^2/\overline{y} > 1$ is satisfied, where $\overline{y}$ and s^2 are, respectively, the sample mean and the sample variance from $\underline{y}$. From (14) and when λ is fixed or known, we easily have by the moment method

$$p^* = \ln[(s^2 - \overline{y})/\lambda]/\ln[\overline{y}/\lambda] = p^*(\lambda). \tag{18}$$

With respect to the unit variance function (1), we can take $\lambda = 1$ in (18) because one stays in the same EDM; after, one can estimate λ in the corresponding $HDM_{p^*}(\mu, \lambda)$.

In order to illustrate (18), we analyze two data sets of Table 3 given by Kokonendji and Khoudar (2004; Table 5.1 and Table 5.2); one of which (Table 5.1) has been revisited by several authors (e.g. Willmot, 1987). This type of data is frequent in marketing, insurance, biometry and financial problems. The data correspond to the number of automobile insurance claims per policy in Germany over 1960 among $n = 23589$ and in Central African Republic over 1984 among $n = 10000$. Both data sets show overdipersion as can be seen in Table 4 with $s^2/\overline{y} \simeq 1.14$. For these data some overdispersed models have been used in the literature among others the negative binomial ($HDM_2 = PTM_2$), strict arcsine (HDM_3) and Poisson-inverse Gaussian (PTM_3) models. It was used a Pearson's chi-square goodness-of-fit statistic and, as observed by several authors, no single probability law seems to emerge as providing "the" best fit.

Table 3: *The two data sets analysed by Kokonendji and Khoudar (2004; Table 5.1 and Table 5.2).*

Data set 1		Data set 2	
No. of claims	No. of policies	No. of claims	No. of policies
0	20592	0	6984
1	2651	1	2452
2	297	2	433
3	41	3	100
4	7	4	26
5	0	5+	5
6	1		

From Table 3 we see that using expression (18) with $\lambda = 1$ we find p^* equal to 2 for data set 1 and 3 for data set 2, leading, respectively, to a negative binomial ($HDM_2 = PTM_2$)

and to a strict arcsine (HDM_3) as models of the Hinde-Demétrio class. As Kokonendji and Khoudar (2004) or Willmot (1987) have shown, the best fit for data set 1 is Poisson-inverse Gaussian model (PTM_3) of the Poisson-Tweedie class with $\chi^2 = 0.48$ for 2 degrees of freedom. For data set 2 the value of χ^2 for the strict arcsine model (HDM_3) is bigger than the value of $\chi^2_2(95\%) = 5.99$, meaning that it does not fit; however, it is the best fit among the models of the Hinde-Demétrio class.

Table 4: *Summary statistics for two data sets of Table 3 [the χ^2 values correspond to the negative binomial and strict arcsine models (Kokonendji and Khoudar, 2004)].*

Data sets	$\bar{y}$	s^2	$s^2/\bar{y}$	p^*	χ^2 values	df
1	0.14	0.16	1.14	1.99	3.12	2
2	0.37	0.42	1.14	3.01	15.64	2

Despite the inefficient estimation of parameter p, in Table 3, we have a good way to obtain the adequate model in the Hinde-Demétrio class or even to think to the Poisson-Tweedie class when $p \simeq 2$. However, inferential techniques are not yet as well developed routinely for these two classes of EDMs. But it should be always interesting to handle simultaneously or separately these two classes, instead for using particular distributions.

Acknowledgments We were partially supported by grants from CNPq, FAPESP and CCInt/USP. The authors would like to thank the Editors and a referee for their helpful comments and suggestions.

References

Castillo, J. and Pérez-Casany, M. (2004). Overdispersed and underdispersed Poisson generalizations. *J. Statist. Plann. Inference*, to appear.

Feller, W. (1943). On a general class of contagious distributions. *Ann. Math. Statist.*, 14 (4), 389-400.

Hinde, J. and Demétrio, C. G. B. (1998). *Overdispersion: Models and Estimation.* São Paulo: ABE.

Hougaard, P., Lee, M-L. T. and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics*, 53, 1225-1238.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions, 2nd ed.* New York: John Wiley & Sons.

Jørgensen, B. (1997). *The Theory of Dispersion Models.* London: Chapman & Hall.

Jourdan, A. and Kokonendji, C. C. (2002). Surdispersion et modèle binomial négatif généralisé. *Rev. Statistique Appliquée*, L (3), 73-86.

Kemp, A. W. and Kemp, C. D. (1966). An alternative derivation of the Hermite distribution. *Biometrika*, 53, 627-628.

Kokonendji, C. C. and Khoudar, M. (2004). On strict arcsine distribution. *Commun. Statist.-Theory Meth.*, 33 (5), 993-1006.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, vol. 1: Models and Applications*, 2nd ed. New York: Wiley.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.*, 18, 1-37.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (eds. J.K. Ghosh and J. Roy), 579-604. Indian Statistical Institute, Calcutta.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial J.*, 2, 113-127.